1990; Do and Hall, 1992), appears to give equally good or better results, can also be used in weighted situations and can also be very effectively smoothed (Hesterberg, 1997).

Finally, Firth and Bennett leave open the question of how to choose x, using an *ad hoc* x in Section 3.2 to estimate the distribution of a sample mean. My approach in bootstrapping is very different—I take the distribution of sample means to be *known* and use sample means of functions of the original data as control variates for estimating the quantiles of general (non-linear) statistics. This is very effective. The distribution of a sample mean is easily and accurately calculated by the first-order Lugannani and Rice saddlepoint (Daniels, 1987), and a single saddlepoint evaluation also estimates the expected value of $c(X)$ if c has the shape of a cumulative distribution function (Hesterberg and Nelson, 1997), e.g. as in logistic regression.

Edward L. Korn and Barry I. Graubard (*National Cancer Institute, Bethesda*)

We are pleased to see that Pfeffermann and his colleagues are pursuing the problem of weighted estimation of variance components when the cluster sample sizes are small (Graubard and Korn, 1996). We noted that unscaled weighted estimators (method C in Graubard and Korn (1996); denoted probability-weighted iterative generalized least squares by Pfeffermann *et al.*) work well with large cluster sample sizes, and we suggested an alternative estimator involving only cluster level weighting for the small sample size case (our method D). Pfeffermann *et al.* suggest two alternative scaled-weighted estimators and tentatively recommend the scaling method 2. To be specific, consider the simplest components-of-variance problem in which a method-of-moments estimator of the within-cluster variance component can be written as

$$\hat{\sigma}^2 = \frac{\sum_{j=1}^{m} w_j \sum_{i=1}^{n_j} w_{i|j}^{*}(y_{ij} - \bar{y}_j)^2}{\sum_{j=1}^{m} w_j \left(\sum_{i=1}^{n_j} w_{i|j}^{*} - 1\right)}.$$

Our estimator involving only cluster level weighting has $w_{i|j}^{*} \equiv 1$ and $\bar{y}_j$ as being the unweighted cluster mean. This estimator is unbiased when the weights are non-informative. Scaling methods 1 and 2 both have $\bar{y}_j$ as the weighted cluster mean, with $w_{i|j}^{*}$ as given by Pfeffermann *et al.* Scaling method 1, proposed by Longford (1995) in the variance component context, also yields an unbiased estimator with non-informative weights. Scaling method 2 is biased with non-informative weights and we therefore cannot recommend it. The simulations provided by Pfeffermann *et al.* and Graubard and Korn (1996) do not really address this issue because the weights are assumed to be constant when they are taken as non-informative, a special case. A direct simulation comparison of Longford's scaling and our cluster-level-only weighting would be interesting.

S. Nadarajah (*University of Plymouth*)

Firth and Bennett are to be congratulated on a very interesting paper. I have two comments.

Now that the UK elections are over it is time to reflect on the predictions of the various polls. The main interest has been on the percentage of votes that each party would expect to receive. I see that the problem of estimating the percentage for a particular party falls nicely under the framework of the paper: $a_i = 100/N$ (where N is the total number of voters interviewed) and y_i is a binary variable indicating whether or not the voter intended to vote for the party. It would be interesting to see how the models discussed in the paper predict the results when fitted appropriately to the data collected from the polls.

The authors discuss several estimates that are design consistent. In sampling techniques it is a common practice that variances of different estimates are compared to give a guide to the practitioner (see for example Cochran (1977)). Have the authors any analytical results that make such comparisons?

J. N. K. Rao (*Carleton University, Ottawa*) **and G. R. Roberts** (*Statistics Canada, Ottawa*)

Hierarchically structured survey data are often analysed using multilevel models. But such analyses often ignore the survey design features, such as unequal probabilities of selection at the first stage of sampling, thus leading to design inconsistent estimators. The paper by Pfeffermann *et al.* makes an important contribution to multilevel analysis by providing, for designs following the underlying model structure, estimators and associated variance estimators that lead to valid design-based inferences. The authors show that we need both the first-stage (level 2) weights w_j and the second-stage (level 1) weights